DA3 Assignment #1

Individual

Deadline: 23.55h Tuesday 16 November 2017

Problem 1.1

Download hotels_all_nov21.csv. Pick a city. Consider hotels and hostels. Consider all with at least 2 stars. You have 7 tasks (1p each):

1. Filter the data to the city of your choice and other characteristics (stars, accomodation type) . Describe the distribution of the price and distance variables. Comment on graphs. (1-2 sentences)
2. Sample definition: You may or may not want to drop some observations; make a choice and argue for it (1-2 sentences).
3. Create a binary variable of distance (below/above cutoff of your choice) and regress price on this binary variable. Report, interpret and visualize the results. (1-2 sentences)
4. Estimate a lowess nonparametric regression of price on distance. Report, interpret and visualize the results. (1-2 sentences)
5. Estimate a simple linear regression of price on distance. Report, interpret and visualize the results. (1-2 sentences)
6. Estimate a linear regression of price on distance that captures potential nonlinearities (polynomials, splies). Report, interpret and visualize the results. (1-2 sentences)
7. Discuss your overall findings. (2-3 sentences)

Problem 1.2: **extra**: for extra 1 point

- See what happens when you estimate your models on a selected sub-sample (ie exclude some hotels based on stars, or location.)
- Discuss the role of cleaning and sample selection.

Submission and grading

- Deadline: 23.55h Tuesday 16 November
  - Late policy: 50% of grade if delay is within 24hs. 0 afterwards.
- Individual
- Max: 7p (+1 extra)
  - Best few submissions will also get an extra point.
    - Best: Good statistics. Clear writing. Nice graphs.
- Submit a single pdf that includes all tables, graphs, text.
- Submit a code.
  - People working in R: you may use R markdown.
- File name: DA3_1_surname_firstname_ddmmyy
  - DA3_1_bekes_gabor_071117.pdf (DA3_1_bekes_gabor_071117.do / .R)